DOCUMENT RESUME

ED 480 069 CG 032 642

AUTHOR	Krug, Samuel E.
TITLE	Maybe We Learned All We Really Needed To Know in Kindergarten: But How Could Anybody Be Sure until We Took the Test?
PUB DATE	2003-08-00
NOTE	13p.; In: Measuring Up: Assessment Issues for Teachers, Counselors, and Administrators; see CG 032 608.
PUB TYPE	Information Analyses (070)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Achievement Tests; Educational Assessment; *Educational Testing; *Learning Processes; *Student Evaluation; *Test Use

ABSTRACT

Tests--typically cognitive achievement tests--are instruments for providing a more sensitive analysis of the learning sequence and of the instruction guiding it. In addition to an overall score, standardized tests provide a variety of subscores that help the teachers identify where a student might be encountering the greatest difficulties. The author argues that testing is not some sort of an optional add-on to education, but is an integral part of the learning process, part of the instruction-test-instruction feedback loop by which students demonstrate knowledge and skill acquisition. (GCP)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 480 069

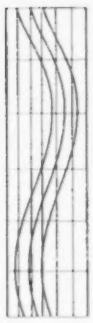
Maybe We Learned All We Really Needed to Know in Kindergarten: But How Could Anybody be Sure Until We Took the Test?

By
Samuel E. Krug

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

CG032642



Chapter 34
Maybe We Learned All We Really Needed to Know in Kindergarten
But How Could Anybody Be Sure Until We Took the Test?

Samuel E. Krug

Let me begin by offering apologies to Robert Fulghum, whose poignant essay I in no way mean to disparage with the title of this chapter. I absolutely agree with him that many lessons first encountered in kindergarten are among life's most significant. I mean only to ask how we know these lessons have been learned. Exposure to the curriculum does not in itself guarantee learning.

Instruction and Testing: The Learning Loop

The problem with measuring learning is that it is an internal process that cannot be directly observed. Kids don't come with indicator lights on their forehead that glow when learning has taken place, although things would certainly be much easier for teachers (and parents) if they did.

By defining learning as a relatively permanent change in behavior, psychology emphasizes the external consequences of the internal process and thereby provides a basis for measuring it. That is, we infer that learning has taken place from measured changes in observable behavior. Put a book in front of most kindergarten students and they are unlikely to be able to tell what it contains other than the pictures. Put that book in front of the same students a few years later and they will be able to retell the story, analyze the characters, and relate the book's contents to situations beyond the story. The book hasn't changed, of course, but the students have. They have learned to read.

That is a test in itself, of course, but a fairly gross one that is most likely insensitive to the many changes that occur in the course of learning to read. Tests—typically cognitive achievement tests—are instruments for providing a more sensitive analysis of the learning sequence and of the instruction guiding it. In addition to an overall score, standardized

tests usually provide a variety of subscores that help the teacher identify where a student might be encountering the greatest difficulties. By reviewing mathematics subscores, for example, a teacher might understand better whether a student's inability to answer multiple-step problems lies in lack of understanding of the basic concepts involved or in weak computational skills. Testing, therefore, is not some sort of an optional add-on to education but is an integral part of the learning process, part of the instruction-test-instruction feedback loop by which students demonstrate knowledge and skill acquisition.

Do Standardized Tests Duplicate What Teachers Are Already Doing?

Why do we need standardized tests? Don't students demonstrate their understanding and skills in numerous ways already? Don't teachers already give a lot of tests that are subsequently reflected in students' grades? This seemed to be the case when we were in school. Aren't standardized tests just duplicating what teachers are already doing at a cost of valuable instructional time?

Teacher-made tests do serve a valuable purpose. Many, perhaps most, are scored and returned in a short time and thus provide rapid feedback to both the teacher and student about how well learning is proceeding. Teachers can tell whether students have understood the material, what amount of repetition is necessary, and how quickly they can move on to other concepts and skills. Because of this formative function, teacher-made tests usually focus on a relatively narrow spectrum of content. That is, they are more likely to focus on content learned over the course of a unit or a chapter than over a year or more of instruction. Teacher tests undoubtedly also serve an important motivational function—and motivation is a critical element of learning—as students try to demonstrate that they have done what the teacher and their parents have asked: They have learned the material.

Classroom tests also have limitations. Perhaps most importantly, they rarely provide insight into performance outside the classroom for the simple reason that no one outside the classroom takes them. Thus they provide a limited range of normative information. Discovering that *nobody* in the classroom knows the answers to *any* of the questions on the test would likely lead a teacher to revisit the subject matter of the test, perhaps trying a different approach to the subject. Finding out that everybody *outside* the classroom knows the answers to *all* the questions would lend a certain urgency to those efforts.

Norms are important because a person is often judged normatively, and the judgments carry important consequences. The Olympic motto *citius, altius, fortius* (swifter, higher, stronger) is normative and, for better or worse, serves as a motto for life in general when resources are limited. Several people may be qualified for a job, for example, but only one, the best qualified, will get the job. Colleges and universities often receive more qualified applications than they can accept. Consequently, knowing where someone is in relationship to everyone else at a time when something can most easily be done about it—the school years—is important.

One challenge teachers face is that they must focus so intently on what goes on inside their classrooms that they don't have the luxury of exploring fully what goes on outside their classrooms. That is to say, their norms are often narrow, limited by necessity to just a few dozen students each year, usually from the same school.

These classroom norms probably reflect community norms pretty well. In days when communities were fairly narrow and isolated, the norms served very well. But the communities we live in today are no longer narrow or isolated, and our schools necessarily must prepare students for much larger communities. Increasingly, the large rather than small communities of which we are members define learning goals and expectations for student performance. Standardized tests that measure the shared learning goals of our larger communities augment the information teachers gain from more localized and focused tests of instructional goals.

Testing and the Standards Movement

Although normative information is valuable, it is not sufficient. Measuring learning against significant criteria others will employ to evaluate performance is critical.

Several years ago I was asked by a school superintendent to help an elementary school's teachers and administrators prepare for a site visit by the state education agency's accreditation team. The new accreditation model emphasized the importance of data in support of the school's assertion that its students' needs were being met. In the process of reviewing the kinds of evidence the school had accumulated, the teachers pointed to a student evaluation form that instructors filled out at the end of the year as guidance for next year's teacher. This seemed to me a sensible item to introduce at the accreditation visit as it appeared to be useful information for maintaining continuity of

instruction as students moved from grade to grade.

After a few minutes of discussion about the form, however, we dropped it from further consideration. It became clear in those moments of discussion that there was no shared understanding of what the information meant. Three teachers at the table explained how they filled out the form and were surprised to discover that they each did it differently. Each applied a different set of standards in completing the form, standards so personal that others were unlikely to interpret the information correctly.

Meeting the standards of the classroom teacher is, of course, critical; however, we need to know that these students also have the skills that society needs to maintain and that they need in order to advance, that they have sufficient knowledge of government, for example, to make them fully participating citizens or that they are sufficiently scientifically and technologically literate to cope with a complex civilization. And the verdict of many important stakeholders in our society is that many—too many—students don't have these skills.

This has directed national attention toward the definition of standards that describe what students are supposed to know and be able to do as a consequence of their education. The standards reflect societal expectations or goals for learning, but they typically also incorporate minimum benchmarks for performance. In Illinois, for example, one of seventeen learning standards for mathematics (see www.isbe.state.il.us/ils/math/mag8.html) states that students will be able to "use algebraic concepts and procedures to represent and solve problems." For students in early elementary grades, this is taken to mean that they will "find the unknown numbers in whole-number addition, subtraction, multiplication, and division situations." For students in late elementary grades this is taken to mean that they will "solve linear equations involving whole numbers." And for 11th- and 12th-grade students this is taken to mean that they will be able to "formulate and solve nonlinear equations and systems, including problems involving inverse variation and exponential and logarithmic growth and decay."

Over the past several years most states have undertaken the development of learning standards in one form or another. A number of broader efforts to develop similar learning standards at a national level have occurred as well. In some areas (e.g., mathematics) significant consistencies in such standards exist across states. In other areas (e.g., social studies) there are important differences among state standards. Such standards are usually developed by educators and the community

in a deliberative process to identify a set of agreed-upon expectations or learning outcomes. As the community broadens, the process becomes more complex.

Nevertheless, once such standards are adopted, the question naturally follows as to how well students are performing in relation to them. This is not a normative function but an evaluative function, and the kinds of tests and items that provide the best normative analysis of student performance are not necessarily the kinds of tests that provide the best evaluative analysis.

As a consequence many, perhaps most, states have undertaken programs to develop criterion-referenced tests that are aligned with their own curriculum standards. In many cases, classroom teachers and curriculum experts from the area are involved in every phase of test development: specifications development, item writing, review, and test assembly. The results of these efforts more often than not evaluate what is thought—at least by public consensus—to represent important ideas and concepts and to do so in psychometrically sophisticated ways.

Tests that teachers construct for use in their own classrooms differ in important ways from these kinds of instruments. Classroom tests are more likely to be narrow in focus, and content is most likely to be covered in a lesson, a unit, or perhaps a semester's learning. These tests are primarily intended to be formative evaluations that provide both the teacher and student with information to guide the instructional process.

In contrast, the content of the state-level test is far more likely to focus on cumulative learning. That is, a state science test administered in the seventh grade doesn't usually assess a specific seventh-grade curriculum but instead assesses things about science that students should have learned in their first seven years of instruction. State-developed criterion-referenced tests are intended to provide summative evaluations for public accountability purposes.

Are the Tests Measuring Real Learning?

One of the criticisms frequently hurled at standardized tests is that they simply measure recall of isolated facts, not true understanding and analysis. The multiple-choice format that remains dominant for most standardized tests is a frequent target for critics who decry its simplistic format.

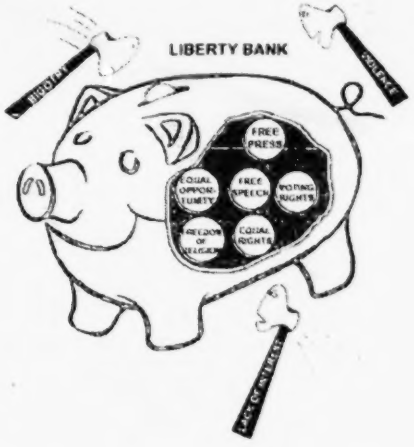
I offer one suggestion in response to such criticism: Take a close look at the test items and student performance on those items. There is

no doubt that some questions that make their way into standardized tests could be improved. Many more enter the test via a process that tends to ensure that surviving items address important issues in important ways.

The reading tests of a generation past, for example, presented students with short, disconnected paragraphs about unusual, often uninteresting topics and asked a narrow range of questions, often factual, about the paragraphs' content. Contemporary reading tests, in contrast, present students with extended texts usually drawn from high-quality contemporary or classic literature. They use informational texts drawn from the kinds of material students are likely to encounter in regular classroom instruction. The questions address complex issues like motivation and character development. They require students to go beyond the text and apply related knowledge to answer questions suggested by the text. "What color was the dragon?" is far less likely to be found in most current reading tests than "What is most likely to happen to the main character when the story ends?"

Contemporary mathematics tests frequently require students to solve multistep problems in order to select the correct answer. Incorrect understanding of the steps to take toward the solution is reflected in incorrect selections.

In social studies, it is not unusual to present elementary students with historically significant political cartoons or archival documents and ask a series of questions that require understanding, analysis, and careful interpretation of the material. Consider this example taken from a fourth-grade Illinois social studies test (Illinois State Board of Education, 2001). Questions that follow the picture ask students what the coins in the bank stand for (rights of Americans) and what the axes stand for (attacks on freedom).



BEST COPY AVAILABLE

The standardized tests of generations past often relied exclusively on multiple-choice items, which emerged as the format of choice during the 1920s. It was attractive to a psychology heavily influenced by behaviorism and the principles of scientific management. In contrast, contemporary tests increasingly rely on constructed-response items that require students to write essays that might take the form of a few sentences summarizing the key concepts of a reading passage. Whereas older "English" tests might have asked students to identify grammatical or stylistic errors in material presented to them through a multiple-choice format, contemporary tests are more likely to ask students to write a two- to three-page essay on presented material. In some cases newer forms of assessment extend to a portfolio of student work accumulated and evaluated over an extended period, although the cost of portfolio assessment has significantly limited its use in large testing programs.

Other Myths About Tests

Despite the facts that testing itself is an integral part of the learning process, that standardized tests supply valuable normative information, and that standards-based tests provide the only credible evidence of whether societally established learning outcomes have been achieved, many still object to what is viewed as an overemphasis on testing in our nation's schools.

The charge is sometimes made that tests narrow the curriculum. This concern usually arises when testing programs of consequence are introduced and teachers begin teaching to the content of those specific tests. This would be a significant criticism if there were, for example, a set of 40 math problems, 20 vocabulary words, 45 historic dates, 20 science facts, or one story that, once taught, would guarantee success on the test. As I have argued earlier, however, most contemporary tests don't rely on simple recall of isolated facts. Students face much more challenging content, content that, quite frankly, deserves to be taught.

In addition, test content changes continually in most testing programs of consequence. Items given one year are unlikely to reappear the next. The item pools for most professionally developed tests are typically extensive. A strategy of teaching students to answer a limited set of test questions, unethical on the surface, would also turn out to be poor strategy in the long run.

Sometimes "teaching to the test" is interpreted as focusing on specific strategies for answering multiple-choice questions. Other times

"teaching to the test" is interpreted as teaching students a formulaic writing style that will ensure high scores by graders. With regard to the former, I would argue that there are at least some strategies involved in answering multiple-choice items that are important life skills, such as carefully considering what the question is asking, evaluating each option before responding, and eliminating the least likely answer to reduce the number of choices when uncertain about a decision. I use those skills every day. I think other people do too. I often think that much of life is a multiple-choice test.

With regard to teaching formulaic writing, the charge extends to the supposition that formulaic writing devalues creativity. Don't get me wrong: I like creativity. But I know that life involves a lot of formulaic writing, so it's not a bad thing to understand how to write in this way. If you disagree, try to get published in a behavioral science research journal an article that deviates too far from the abstract-introduction-method-results-discussion format of the *Publication Manual of the American Psychological Association*. Much correspondence required in business follows standard forms as well. So formulas aren't all that bad. They are certainly better than no writing or writing in which it is impossible to detect an orderly presentation of ideas.

The charge is sometimes made that testing takes time that would otherwise be used for teaching and instruction. The usual targets of the criticism are the large state accountability programs, and the implication is that these ponderous programs consume hundreds of hours of valuable instruction time. These programs don't involve hundreds of hours of testing time; at least I haven't encountered one yet that did. In Illinois, for example, the state programs make the greatest demand on student time in 11th grade, when the Prairie State Achievement Test requires about seven and a half hours spread over two days. Almost half of that time is devoted to taking a college entrance examination that the majority of the students would have taken anyway. At other grades, state testing requires no more than five or six hours. Across the nine months students attend school, that doesn't seem unreasonably burdensome.

The charge is sometimes made that testing instills competitiveness. Are we to believe that people were not competitive before educational testing was invented? At the time this chapter was written, the 19th winter Olympic games had just ended at Salt Lake. Talk about competition. Competition is a fact of human nature engendered by the economic reality that resources are limited. Testing may be an unwelcome reminder of that economic reality, but it is not the cause of

competition.

A related charge is sometimes made that testing demoralizes students. It is human nature to feel bad when we do poorly. But to blame the test is very much like killing the messenger. The federal education legislation that requires states to conduct assessments of all students in their public schools also requires the reporting of student results in discrete performance categories. One or more of these categories is usually undesirable. Nevertheless, the intent of that legislation is to establish clear goals for improvement and document adequate yearly progress toward moving students out of undesirable categories.

Although it would be preferable to be reinforcing rather than demoralizing and to give out only good news, if the reality is otherwise, then you can't always give good news. During the mid-twentieth century there was substantial belief in the efficacy of our system of public instruction. For most of the first half of that century the public schools of our great cities were some of the best of that system. Although we awoke in 1984 to find that we were a nation at risk, the system did not fail overnight. To a large extent, the current test culture is a consequence of too much good news—undocumented news—for too long. The public requires some assurance that 1984 and the decades of indifference that preceded 1984 won't happen again, and tests provide some measure of that assurance.

The charge is sometimes made that we are testing students too early, that students in the early grades are too young to be other than dismayed by tests. Most state accountability programs require testing for the first time at third grade or later. The No Child Left Behind Act of 2001 requires annual testing of students, but only from third through eighth grades. The problem for me is not testing too early but too late. By the third grade, some of the most significant learning students encounter has or should have taken place. Students in the first two or three years of elementary school, for example, learn how to read. After that they will, for the most part, read to learn, but only if the lessons of the first critical year or two are learned. If not, they will play catch-up for most of the rest of their academic careers, often unsuccessfully. If the first time we are aware that students are being left behind is at the end of third grade, some crucial opportunities have already vanished and can only be made up with considerable effort.

The charge is sometimes made that so much testing of students amounts to little more than weighing the pig over and over. That is, there is little value in testing repeatedly because the action itself does

nothing to increase achievement. The criticism misses the essential point that obtaining a measurement is not the purpose. Instead the purpose of taking the measurement is to act on the information it provides.

There are many who agree that public education is in a critical state, that too many students are being denied the quality of instruction to which they are entitled, and that our society is unlikely to continue advancing without significant intervention. In the eyes of these people tests provide the continual monitoring needed to ensure that the student continues to make progress.

The charge is sometimes made that tests have become the ultimate criterion for obtaining diplomas or other valued credentials. That is, the stakes associated with the tests are just too high. At the high school level, the number of states that have introduced a testing requirement for granting a degree has increased in the last few years. But as Cizek (2001) correctly observed, the testing requirement is not an ultimate criterion, just the latest one. There were already a number of criteria in place that students had to satisfy to receive a diploma. These requirements addressed the number of credit hours students must accumulate, completion of specific course requirements (e.g., consumer education, physical education, government), and attendance requirements. The test requirement is one that must be met in addition to these others, but failure to comply with the others denies a diploma just as quickly as a poor score on the test does. Moreover, diploma tests can be attempted more than once, so if a student has trouble with one administration he or she has other opportunities.

The introduction of a testing requirement is little more than what has been done for many years in the area of professional certification and licensure. A person may spend years in medical school or law school and meet the moral character requirements for practice, but absent a passing score on the licensure test, this person will not practice in the profession. Rather than demeaning the four years of instruction that lead to graduation, it seems more likely that the introduction of a testing requirement into the process will result in increased perceived value of the credential as it seems to have done for many professions.

On the other hand, the charge is almost never made that tests are not used enough, but I think that is often the case, at least with respect to their results. More often than not, institutions (e.g., state departments, districts, schools) spend far more time and effort administering test programs than they do studying their results.

Despite the various criticisms that have been directed toward educational testing, there appears to be strong public support for it. If

there were not, it would have been far more difficult to achieve passage of the No Child Left Behind Act of 2001, which mandates testing in every grade. The passage of that legislation, and the failure of the previous administration's legislation, which proposed testing all students at just two grades suggest that support has actually increased in the last few years.

This support most likely arises from a pervasive belief that standardized tests are ultimately among the fairest and most accurate indicators of the condition of educational achievement available to us. The classroom tests that teachers administer and the grades they derive from them serve an important function, but they don't always give as clear a picture of performance beyond the classroom as we require.

Despite the many objections, the fact remains that standardized testing itself is an integral part of the learning process, that these tests supply valuable normative information, and that criterion-referenced tests provide the credible evidence of whether societally established learning outcomes—including all we are supposed to have learned in kindergarten—have been achieved.

References

Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice, 20*(4), 19–27.

Illinois State Board of Education. (2001). *Illinois Standards Achievement Test sample social science materials.* Springfield: Author.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
X	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").